The Linear Model From A Coordinate-Free Viewpoint*

BU-~~238~~ 240-M Robert Jacobsen March, 1967

Abstract

This paper uses the coordinate-free approach to linear algebra to simplify and unify the explanation of statistical inference in the linear model.

* Paper No. BU-~~238~~ 240-M in the Biometrics Unit, Plant Breeding and Biometry Department, Cornell University, Ithaca, New York.

The Linear Model From A Coordinate-Free Viewpoint*

BU-~~238~~ 240-M   Robert Jacobsen   March, 1967

I. Some Concepts in Linear Algebra

1. Let V be a vector space, V_1, V_2 subspaces of V. Then V is <u>the sum</u> of V_1 and V_2, written $V = V_1 + V_2$, if each vector $v \in V$ can be written as $v_1 + v_2$, where $v_1 \in V_1$, $v_2 \in V_2$. If this decomposition is unique, then we write $V = V_1 \oplus V_2$, and call V the <u>direct sum</u> of V_1 and V_2.

A necessary and sufficient condition for $V_1 + V_2$ to be a direct sum is that $V_1 \cap V_2 = 0$, i.e., the subspaces must be linearly independent.

<u>Example</u>: R^3 is the direct sum of V_1, the span of $(1,1,1)$ and V_2, any plane through the origin not containing $(1,1,1)$.

2. Let V be a real inner product space. An <u>inner product</u> $(\cdot,\cdot)$ is a bilinear, symmetric function on $V \times V$ s. t. $(x, x) \geq 0$ and $(x, x) = 0$ iff $x = 0$.

3. Let V_1 be a subspace of V. <u>The orthogonal complement of</u> V_1, denoted $V_1^{\perp}$ (read V_1 perp), is the set of all $v \in V$ s. t. $v \perp v_1$ for all $v_1 \in V_1$ ($v \perp v_1$ means $(v_1, v_1) = 0$). $V_1^{\perp}$ is a unique subspace. And $V_1^{\perp} \cap V_1 = 0$, since $x \in V_1^{\perp} \cap V_1 = 0$, since $x \in V_1^{\perp} \cap V_1 \Rightarrow (x, x) = 0 \Rightarrow x = 0$.

---

* Paper No. BU-~~238~~ 240-M in the Biometrics Unit, Plant Breeding and Biometry Department, Cornell University, Ithaca, New York.

4. If V_1 is a subspace of V, and V is finite dimensional, then $V = V_1 \oplus V_1^{\perp}$. The above is called an <u>orthogonal direct sum decomposition</u> of V.

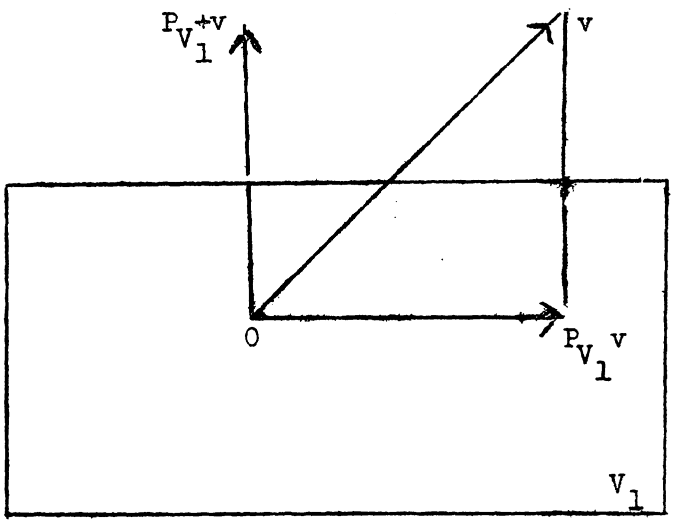
<u>Proof</u>:

Pick an orthonormal basis $\xi_1, \cdots, \xi_m$ for V_1.
Let $v \in V$, $v_1 = \sum_{i=1}^{m} (v, \xi_i)\, \xi_i$. Then $v_1 \in V_1$. Let $x = v - v_1$.

Then $(x, \xi_j) = (v, \xi_j) - (v_1, \xi_j) = (v, \xi_j) - (\sum_i (v, \xi_i)\, \xi_i, \xi_j)$

$= (v, \xi_j) - (v, \xi_j) = 0 \qquad j = 1, \cdots, m.$ So $x \in V^{\perp}$.
So $V = V_1 + V_1^{\perp}$. But $V_1^{\perp} \cap V_1 = 0$. So $V = V_1 \oplus V_1^{\perp}$.

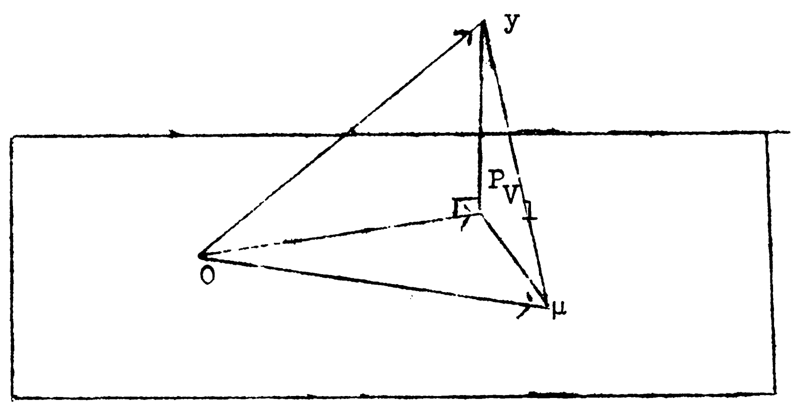
5. Let V_1 be a subspace of V. Decompose $V = V_1 \oplus V_1^{\perp}$. If $v \in V$, $v = v_1 + v_1^{\perp}$. Define $P_{V_1} v = v_1$. The map P_{V_1}, which is linear, is called <u>the orthogonal projection on</u> V_1. Note that $I - P_{V_1}$ is <u>the orthogonal projection on</u> $V_1^{\perp}$, for $(I - P_{V_1})\, v = v - P_{V_1} v = v_1^{\perp}$.



Note that a projection is <u>symmetric</u> and <u>idempotent</u>. For, let $x, y \in V$. $x = x_1 + x_1^{\perp}$, $y = y_1 + y_1^{\perp}$, the decomposition of x and y induced by $V = V_1 \oplus V_1^{\perp}$. Then $(x, P_{V_1} y) = (P'_{V_1} x, y)$ defines the <u>adjoint</u> P'_{V_1} of P_{V_1}. (See 9). But l.h.s. $= (x_1 + x_1^{\perp}, y_1) = (x_1, y_1) = (x_1, y_1 + y_1^{\perp}) = (P_{V_1} x, y)$. So $P_{V_1} = P'_{V_1}$.

<u>Idempotency</u> of P_{V_1} follows, as $P^2_{V_1} x = P_{V_1} P_{V_1} x = P_{V_1} x_1 = P_{V_1} x$.

Projection has a further important property. Given $v \in V$, $P_{V_1} v$ is the point in V_1 which is closest to v in the inner product distance.



<u>Proof</u>:

Let $\mu \in V_1$

$y - \mu = (y - P_{V_1} y) + P_{V_1} y - \mu)$

$P_{V_1} y - \mu \in V_1$. $y - P_{V_1} y \in V_1^{\perp}$.

Hence, their lengths² add.

$\|y - \mu\|^2 = \|y - P_{V_1} y\|^2 + \|P_{V_1} y - \mu\|^2$.

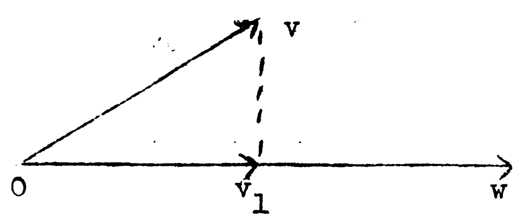
So $\inf\limits_{\mu \in V_1} \|y - \mu\|^2$ occurs for $\mu = P_{V_1} y$.

6. <u>Interpretation of the inner product</u>

Let $v, w \in V$, V_1 = span of w. Then $(v, w) = \|P_{V_1} v\| \; \|w\|$, that is, the inner product of two vectors is the length of the projection of one (say v) on the other times the length of the vector projected onto (w).

<u>Proof</u>:

Decompose $V = V_1 \oplus V_1^{+}$. Then $P_{V_1} v = v_1$ and $\|P_{V_1} v\| = \|v_1\|$. Now $w = cv_1$, and $\|w\| = c\,\|v_1\|$. But $\underline{(v, w)} = (v_1 + v_1^{+}, w) = (v_1, w) = (v_1, cv_1)$ $= c\,\|v_1\|^2 = \|v_1\| \cdot c\,\|v_1\| = \underline{\|P_{V_1} v\| \cdot \|w\|}$.



7. The dual space of V, denoted V', is the vector space of all linear functionals on V.

If V is an inner product space, and $v \in V$, $(\cdot, v)$ is a l. functional on V. The space S of such funct'ls is isomorphic to V. But, if V is <u>finite-dimensional</u>, V is isomorphic to V'. Hence S is isomorphic to V and the *dual space of* V *is simply all funct'ls of the form* $(\cdot, v)$, $v \in V$, in the finite-dimensional case.

8. Given a subspace W of V, *the annihilator of W*, denoted W^0, is the set of l. funct'ls on V that annihilate W, i.e., s. t., if $w^0 \in W^0$, then $w^0(w) = 0$, $\forall\, w \in W$.

If $w^0 \in W^0$, then w^0 assigns the same value to all points of V in the cosets $v \oplus W$. Hence w^0 can be regarded as a l. funct'l on the *quotient space* V/W (read V mod W). Or, writing $V = W \dotplus W^+$, $w^0 \in (W^+)'$. Now w^0 has the form $(\cdot, v)$. So $v \in W^+$. We thus conclude that a *l. funct'l* $\underline{f}$ *is in* S', where S is a subspace of V iff it annihilates S^+ iff $\underline{f} \in S$. ($f(\cdot) = (\cdot, \underline{f})$). Conversely, given $f(v) = (v, \underline{f})$, let $W = (\text{span of } \underline{f})^+$. Then $f \in W^0$ and $f \in (W^+)' = (\text{span of } \underline{f})'$.

9. Let A: $V \to W$ be linear, V, W finite-dimensional inner product spaces. Then A': $W \to V$, the *adjoint* of A, is defined by $(x, Ay) = (A'x, y)$, $\forall\, x \in W$, $y \in V$. This definition is justified as follows:

A' arises as $\psi^{-1} \circ M \circ \psi$, where ψ is the cononical 1 - 1 map from a space to its dual, and M is the unique map on $W' \to V'$ induced by A by

$$(M(w'))(v) = w'\,(A(v)), \ \forall\, w' \in W',\ v \in V.$$

Picture:
$$\begin{tikzcd} V' & W' \arrow[l, "M"'] \\ V \arrow[r, "A"] & W \end{tikzcd}$$

For, let $x \in W$, $y \in V$. Then $(A'x, y) = (\psi^{-1} \circ M \circ \psi(x), y)$

$$= (M \circ \psi(x))\,(y) = (\psi(x))\,(Ay) = (x, Ay).$$

If A: $V \to V$ and $A = A'$, we say $\underline{A}$ $\underline{\text{is}}$ $\underline{\text{symmetric}}$.

10. Let V be a finite-dim. inner product space. Let f be a bilinear function on $V \times V \to R'$. Then $\exists\, A$: $V \to V$ linear s. t. $f(x, y) = (x, Ay)$. If f is symmetric, $A = A'$.

Proof:

Let $\xi_1, \cdots, \xi_n$ be an orthonormal basis for V.
Define $A(\xi_j) = \sum\limits_i f(\xi_i, \xi_j)\, \xi_i$. Then $(\xi_i, A\,\xi_j) = f(\xi_i, \xi_j)$. Now, let $x, y \in V$, $x = \Sigma\, x_i \xi_i$, $y = \Sigma\, y_j \xi_j$. Then $f(x, y) = f(\Sigma\, x_i \xi_i, \Sigma\, y_j \xi_j)$
$= \Sigma\, x_i\, f(\xi_i, \Sigma\, y_j \xi_j) = \sum\limits_i x_i \sum\limits_j y_j\, f(\xi_i, \xi_j) = \sum\limits_i x_i \sum\limits_j y_j\, (\xi_i, A\,\xi_j)$
$= \sum\limits_i x_i\, (\xi_i, Ay) = (x, Ay)$.
If $f(x, y) = f(y, x)$, $(x, Ay) = (Ax, y) = (A'x, y)$. So $A = A'$.

11. Let A: $V \to W$ be linear, V, W vector spaces. Then

$\underline{\text{Ker } A} = \{v \in V:\ A(v) = 0\}$, and

$\underline{\text{Im } A} = \{w \in W:\ w = Av \text{ for some } v \in V\}$.

12. Let A: $V \to W$ be linear. Then $\text{Ker } A = (\text{Im } A')^{\perp}$ and $\text{Im } A = (\text{Ker } A')^{\perp}$.

Proof:

Let $v \in \text{Ker } A$. Then, for any $w \in W$, $(v, A'w) = (Av, w) = (0, w) = 0$. Conversely, if $(v, A'w) = (Av, w) = 0$ for all $w \in W$, then $Av = 0$, so $v \in \text{Ker } A$. The second part follows similarly.

Corollary:

Let $A: V \to V$ be linear and summetric. Then $\text{Ker } A = (\text{Im } A')^{+}$ $= (\text{Im } A)^{+}$. Thus, a symmetric map induces a natural orthogonal decomposition of the domain $V = \text{Ker } A \oplus \text{Im } A$.

II. The Concept of Random Point

We consider a sample point $\mathfrak{y}$, which ranges over an n-dimensional inner product space V. The inner product is denoted by $(\cdot,\cdot)$. To develop the formalism, assume that (Ω, β, P) is a probability triple.

Definition:

$\mathfrak{y}: \Omega \to V$ is called a random point if $\{\omega \in \Omega \mid (v, \mathfrak{y}(\omega)) \leq c\} \in \beta$, $\forall v \in V$, $\forall c \in R'$.

We are saying that all linear functionals of $\mathfrak{y}$ must be measurable, so that the coordinates of $\mathfrak{y}$ with respect to any origin and basis are random variables.

Definition:

$E\mathfrak{y}$, the expectation of $\mathfrak{y}$, is defined to be the unique point $\mu \in V$ s. t. $E^{*}(v, \mathfrak{y}) = (v, \mu)$, $\forall v \in V$, provided $(v, \mathfrak{y})$ is in $L_1(\Omega,\beta,P)$, $\forall v \in V$. (E^{*} denotes ordinary expectation).

Proof:

Such a unique point exists, as $E^*(\cdot, \mathcal{Y})$ is a linear functional on a finite-dimensional space.

Definition:

$\operatorname{Cov}\mathcal{Y}$, the covariance of $\mathcal{Y}$, is the unique linear map on V to V s. t. $(x, \operatorname{Cov}\mathcal{Y}\, z) = \operatorname{Cov}[(x, \mathcal{Y}), (z, \mathcal{Y})] \quad \forall\, x, z \in V$, provided $(v, \mathcal{Y})$ is in $L_2\,(\Omega, \beta, P)$, $\forall\, v \in V$.

Proof:

$(x, \mathcal{Y}), (z, \mathcal{Y}) \in L_2 \Rightarrow |\operatorname{Cov}[(x, \mathcal{Y}), (z, \mathcal{Y})] \leq \operatorname{Var}(x, \mathcal{Y}) \operatorname{Var}(z, \mathcal{Y}) < \infty$. So the r. h. s. exists and is finite. Further, the map from $V \times V$ to R^1 by $[x, z] \to \operatorname{Cov}[(x, \mathcal{Y}), (z, \mathcal{Y})]$ is bilinear and symmetric. Hence, it has a unique representation on V, which is symmetric $(\operatorname{Cov}\mathcal{Y} = (\operatorname{Cov}\mathcal{Y})')$.

Lemma:

Let $A: V \to U$ be linear. Then $E(A\mathcal{Y}) = AE\mathcal{Y}$, provided $E\mathcal{Y}$ exists.

Proof:

We want to show, for all $u \in U$, that

(1) $$E^*(A\mathcal{Y}, u) = (AE\mathcal{Y}, u).$$

Now, $E^*(A\mathcal{Y}, u) - (AE\mathcal{Y}, u) = E^*[(A\mathcal{Y}, u) - (AE\mathcal{Y}, u)]$

(2) $$= E^*[(A\mathcal{Y} - AE\mathcal{Y}, u)].$$

But $A\mathcal{Y} - AE\mathcal{Y} = A(\mathcal{Y} - E\mathcal{Y})$ and so $(A\mathcal{Y} - AE\mathcal{Y}, u) = (A(\mathcal{Y} - E\mathcal{Y}), u)$
$= (\mathcal{Y} - E\mathcal{Y}, A'u)$ Call $A'u = v$.
Then (2) becomes $E^*[(\mathcal{Y}, v) - (E\mathcal{Y}, v)] = E^*(\mathcal{Y}, v) - (E\mathcal{Y}, v)$
$= 0$, for all $v \in \operatorname{Im} A' = V$.

So (1) holds, for all $u \in U$.

<u>Lemma</u>:

Let $A: V \to U$ be linear.

Then $\operatorname{Cov} A\mathcal{Y} = A \operatorname{Cov}\mathcal{Y} A'$.

<u>Proof</u>:

$\operatorname{Cov} A\mathcal{Y}$ is the unique l.t. on $U \to U$ s. t. $\operatorname{Cov}[(x, A\mathcal{Y}), (x', A\mathcal{Y})]$
$= (x, \operatorname{Cov} A\mathcal{Y} x')$, $\forall\, x, x' \in U$.
Now $(x, (A \operatorname{Cov}\mathcal{Y} A')x') = (A'x, \operatorname{Cov}\mathcal{Y}(A'x')) = \operatorname{Cov}[(A'x, \mathcal{Y}), (A'x'\mathcal{Y})]$
$= \operatorname{Cov}[(x, A\mathcal{Y}), (x', A\mathcal{Y})]$, $\forall\, x, x' \in U$.

We now give a geometric meaning to <u>singular</u> <u>distributions</u>.

<u>Lemma</u>:

Let $\mathcal{Y}$ be a random point in V. Suppose $\operatorname{Cov}\mathcal{Y} = A$, $\operatorname{Im} A = \Omega$, $\operatorname{Ker} A = N$.
Then $\mathcal{Y} \in \Omega$ wpr 1, and Ω is the smallest such linear subspace.

<u>Proof</u>:

By the definition of $\operatorname{Cov}\mathcal{Y}$,

$$\operatorname{Cov}[(x, \mathcal{Y}), (z, \mathcal{Y})] = (x, Az),$$

$\forall\; x, z \in V$.

Let $z \in \operatorname{Ker} A$.

Thus, $\{v: (v, \mathcal{Y}) = 0 \text{ wpr } 1\} = (\mathrm{Im}\, A)^{\perp}$ and so $\mathcal{Y} \in (\mathrm{Im}\, A)^{\perp\perp} = \mathrm{Im}\, A$ wpr 1 and it is the smallest such linear subspace.

Then, $(x, Az) = (x, 0) = 0, \quad \forall\, x \in V$.

So $\mathrm{Cov}\,[(z, \mathcal{Y}), (z, \mathcal{Y})] = \mathrm{Var}\,(z, \mathcal{Y}) = 0$.

So $(z, \mathcal{Y}) = 0$ wpr 1.

But, since $A = A'$ and, for a symmetric map, $\mathrm{Ker}\, A = (\mathrm{Im}\, A')^{\perp}$ $= (\mathrm{Im}\, A)^{\perp}$, we have shown that

$$(\mathrm{Im}\, A)^{\perp} \subset \{v: (v, \mathcal{Y}) = 0 \text{ wpr } 1\}.$$

Conversely, let v be s. t. $(v, \mathcal{Y}) = 0$ wpr 1.

Then $0 = \mathrm{Cov}\,[(x, \mathcal{Y}), (v, \mathcal{Y})] = (x, Av), \quad \forall x \in V$. So $v \in \mathrm{Ker}\, A = (\mathrm{Im}\, A)^{\perp}$.

<u>Lemma</u>:

Let $A: V \to W$

$B: V \to U$

be linear maps. Then $A\mathcal{Y}$ and $B\mathcal{Y}$ are uncorrelated iff $A \,\mathrm{Cov}\, \mathcal{Y}\, B' = 0$.

<u>Proof</u>:

By <u>defn</u>., $A\mathcal{Y}$ and $B\mathcal{Y}$ are uncorrelated iff $\mathrm{Cov}\,[(w, A\mathcal{Y}), (u, B\mathcal{Y})]$ $= 0, \quad \forall\, w \in W,\ u \in U$. But, $(w, A\mathcal{Y}) = (A'w, \mathcal{Y})$, $(u, B\mathcal{Y}) = (B'u, \mathcal{Y})$.

So $\mathrm{Cov}\,[(w, A\mathcal{Y}), (u, B\mathcal{Y})] = (A'w, (\mathrm{Cov}\,\mathcal{Y})\, B'u) = (w, (A \,\mathrm{Cov}\, \mathcal{Y}\, B')u) = 0$, $\forall\ w, u$ iff $A \,\mathrm{Cov}\, \mathcal{Y}\, B' = 0$.

Corollary:

Suppose $A: V \to V$, $B: V \to V$ symmetric and $\mathrm{Cov}\,\mathcal{Y} = \sigma^2 I$. Then $A\mathcal{Y}$ and $B\mathcal{Y}$ are uncorrelated iff Im A is orthogonal to Im B, i.e., $A\mathcal{Y} + B\mathcal{Y}$ wpr 1.

Proof:

$$(Ax, By) = (x, A'By) = 0 \quad \forall\ x, y \in V \text{ iff } A'B = AB' = 0.$$

III. The Standard Linear Model and B.L.U.E. Estimation

1. Definition:

$\mathcal{Y}$ is called weakly spherical if $\mathrm{Cov}\,\mathcal{Y} = \lambda I$ for some $\lambda > 0$, where I denotes the identity map.

The standard linear model assumptions are that $\mathcal{Y}$ is a random point in V, an n-dimensional inner product space, and $\mathrm{Cov}\,\mathcal{Y} = \sigma^2 I$, $E\,\mathcal{Y} = \mu$, where $\mu \in \Omega$, a given r-dimensional subspace of V, called the mean space. V is called the observation space. The mean space is described as the image space of a linear map $X: \Theta \to V$, where Θ is a p-dim. space, called the parameter space.

There are, in general, many maps describing the same image space Ω. X is not assumed to be 1 - 1, in general.

Consider, for the moment, that X is not necessarily linear and not 1 - 1. Suppose we want to estimate the "true" parameter point, θ. Even if we ascertained $\mu = E$ exactly, we would still be unable to tell which of the parameter points in $X^{-1}\{\mu\}$ was the "true" one. Thus, the only parametric functions that are identified (estimable) are functions of μ. In particular, assume then, that X is linear.

Definition:

A linear parametric functional (l.p.f.) f on Θ to R^1 is called estimable (est'le) if it depends on θ only through μ, i.e., there exists a linear functional g on Ω s. t. the diagram below commutes.

$$\begin{tikzcd} R^1 & & \Omega \arrow[ll, "g"'] \\ & \Theta \arrow[ul, "f"] \arrow[ur, "X"'] & \end{tikzcd}$$

Now $g \in \Omega'$, the dual of Ω, implies that $\underline{g} \in \Omega$. Let $e \in \Omega^{+}$, $x = \underline{g} + e$. Then $(x, \mu) = (\underline{g} + e, \mu) = g(\mu)$, $\forall\ \mu \in \Omega$. So, the coefficient vector in this representation of an est'le l.p.f. is only unique mod Ω^{+}, and any est'le l.p.f. can be abbreviated (x, μ), $\mu \in \Omega$. Since any est'le l.p.f. f must be constant on $X^{-1}\{\mu\}$, i.e., on the cosets $\theta \oplus \operatorname{Ker} X$, $\theta \in \oplus$, f must annihilate $\operatorname{Ker} X$. So f is est'le iff $\underline{f} \in (\operatorname{Ker} X)^{+} = \operatorname{Im} X'$.

Definition:

A linear estimator (est'r) is any linear functional on V, composed with . A vector linear est'r is any linear map on V to W, composed with $\mathcal{Y}$.

Definition:

An est'r $h(\mathcal{Y})$ is called an unbiased est'r of a parametric function f if $E_\theta\, h(\mathcal{Y}) = h(X(\theta)) = f(\theta), \quad \forall\, \theta \in \Theta$.

B.l.u.e. Estimation

Now, consider the estimation of an est'le l.p.f. (x, μ), $x \in V$, by a linear est'r $(h, \mathcal{Y})$, $h \in V$. If loss is squared error, then risk is bias^2 + variance. Now $\mathrm{Var}_\theta\,(h, \mathcal{Y}) = (h, \sigma^2 Ih) = \sigma^2\|h\|^2$, which is independent of θ. And $\mathrm{bias}^2_\theta = [E_\theta\,(h, \mathcal{Y}) - (x, X\theta)]^2 = (h-x, X)]^2 = (h-x, X\theta)^2 = (X'(h-x), \theta)^2$, which, for any fixed h, x s. t. h-x is not in Ker $X' = (\mathrm{Im}\ X)^+ = \Omega^+$, grows without bound, as $\|\theta\| \to \infty$ along the line determined by $X'(h-x)$.
Now $h = x + h-x$ and $(h, \mathcal{Y}) = (x, \mathcal{Y}) + (h-x, \mathcal{Y})$. So $E(h, \mathcal{Y}) = (x, \mu) + (h-x, \mu) = (x, \mu)$ for all $\mu \in \Omega$ iff $h-x \in \Omega^+$.
Hence, the minimax linear est'r (linear est'r whose maximum risk is smallest, if it exists) will lie in the class of unbiased linear est'rs of (x, μ), i.e., those s. t. $h-x \in \Omega^+$. But, for an unbiased est'r, the risk is just the variance. Now, $h - P_\Omega x = h-x + P_{\Omega^+} x$, which is therefore in Ω^+, if (h,) is unbiased. So $h - P_\Omega x + P_\Omega x$ and $\mathrm{Var}\,(h, \mathcal{Y}) = \sigma^2\| h - P_\Omega x + P_\Omega x \|^2 = \sigma^2\,\{\|h - P_\Omega\, x\|^2 + \|P_\Omega\, x\|^2\}$. This is minimized over h s. t. $h-x \in \Omega^+$ by choosing $h = P_\Omega\, x$. Thus, the minimax linear est'r of (x, μ) is $(P_\Omega\, x, \mathcal{Y}) = (x, P_\Omega \mathcal{Y})$. It is unbiased and is, in fact, the best linear unbiased est'r (b.l.u.e.) of (x, μ) (best means smallest variance in the class of unbiased linear est'rs). Since, if $h \in \Omega$, $(h, \mathcal{Y})$ is the b.l.u.e. of its expectation (h, μ), the class of b.l.u.e.s is precisely the class of est'rs of the form $(h, \mathcal{Y})$, where $h \in \Omega$.

Notice that b.l.u.e. estimation is linear. If the b.l.u.e. of (x_1, μ) is $(h_1, \mathcal{Y})$, and the b.l.u.e. of (x_2, μ) is $(h_2, \mathcal{Y})$, the b.l.u.e. of $(x_1, \mu) + (x_2, \mu) = (x_1 + x_2, \mu)$ is $(x_1 + x_2, P_\Omega \mathcal{Y}) = (x_1, P_\Omega \mathcal{Y}) + (x_2, P_\Omega \mathcal{Y}) = (h_1, \mathcal{Y}) + (h_2, \mathcal{Y})$.

We will call $P_\Omega \mathcal{Y}$ the vector b.l.u.e. of μ. For, $P_\Omega \mathcal{Y}$ is linear, $E\, P_\Omega \mathcal{Y} = P_\Omega\, E\, \mathcal{Y} = \mu$. And $P_\Omega \mathcal{Y}$ is the vector l.u.e. that leads to minimum variance for all derived est'rs of linear functionals of μ.

<u>Lemma</u>:

If D: $V \to V$ is linear, $D \neq P_\Omega$ and $E_\theta D \mathcal{Y} = X\,\theta$, $\forall \theta \in \Theta$, then, for all $x \in \Omega$, $\mathrm{Var}\ (x, P_\Omega \mathcal{Y}) < \mathrm{Var}\ (x, D\,\mathcal{Y})$.

<u>Proof</u>:

$D = D - P_\Omega + P_\Omega.$

$D' = D' - P_\Omega + P_\Omega.$

$E\, D\, \mathcal{Y} = D\,\mu = (D - P_\Omega + P_\Omega)\,\mu$

$= (D - P_\Omega)\,\mu + \mu = \mu, \quad \forall\ \mu \in \Omega$ iff $\mathrm{Ker}\ D - P_\Omega \supset \Omega$.

Now $(x, D\,\mathcal{Y}) = (x, (D - P_\Omega)\,\mathcal{Y}) + (x, P_\Omega \mathcal{Y})$. But $(D - P_\Omega)\,P_\Omega'$ $= (D - P_\Omega)\,P_\Omega = 0$, as $\mathrm{Ker}\ D - P_\Omega \supset \Omega$. Hence, $(D - P_\Omega)\,\mathcal{Y}$ and P_Ω are uncorrelated and $\mathrm{Var}\ (x, D\,\mathcal{Y}) = \mathrm{Var}\ (x, (D - P_\Omega)\,\mathcal{Y} + \mathrm{Var}\ (x, P_\Omega \mathcal{Y})$.

Now $\mathrm{Ker}\ D' - P_\Omega = \mathrm{Ker}\ (D - P_\Omega)' = (\mathrm{Ker}\ D - P_\Omega)^\perp \subset \Omega^\perp$. So $\mathrm{Var}\ (x, (D - P_\Omega)\,\mathcal{Y})$ $= \mathrm{Var}\left((D - P_\Omega)'\, x, \mathcal{Y}\right) = \sigma^2 \|(D' - P_\Omega)\, x\|^2 > 0, \quad \forall\ x \in \Omega.$

Least-squares Estimation

The connection between b.l.u.e. estimation and least squares estimation is that they are equivalent. For, consider the problem of least squares. For each realization y of $\mathcal{Y}$, we want to find the point μ^* in Ω which is closest to y in the inner product distance. But this $\mu^* = P_\Omega y$ = the b.l.u.e. estimate of μ.

<u>Definition</u>:

Let $\hat{\Theta}(y)$ be any point in $X^{-1}\{P_\Omega y\}$. $\hat{\Theta}(\mathcal{Y})$ is called a <u>least squares est'r</u> of the parameter point θ. If f is an est'le l.p.f., we define $f(\hat{\Theta}(\mathcal{Y}))$ to be a <u>least squares est'r of f</u>.

But the diagram

$$\begin{tikzcd} R^1 & & \Omega \arrow[ll, "g"'] \\ & \Theta \arrow[ul, "f"] \arrow[ur, "X"'] & \end{tikzcd}$$

shows that $f(\hat{\Theta}(\mathcal{Y})) = g(P_\Omega \mathcal{Y})$ = the b.l.u.e. of f.

Hence, the <u>Gauss-Markov theorem</u> is proved: the b.l.u.e of f, an est'le l.p.f. is given be $f(\hat{\Theta}(\mathcal{Y}))$, where $\hat{\Theta}(\mathcal{Y})$ is any least-squares est'r of θ.

Useful fact in computing projections

The fact that $(x, P_\Omega \mathcal{Y})$ is the b.l.u.e. of (x, μ) gives us an easy way to find P_Ω y in some situations. Let $\xi_1, \cdots, \xi_n$ be an orthonormal basis for V. Then $(\xi_i, P_\Omega \mathcal{Y})$, the i^{th} coordinate of $P_\Omega \mathcal{Y}$ w.r.t. the ξ's is the b.l.u.e. of (ξ_i, μ), the i^{th} coordinate of μ (mean of the i^{th} cell).

Suppose we have a spanning set for the b.l.u.e.s, i.e., vectors t_i, $i = 1, \cdots, \dim. \Omega$, spanning Ω. Then $E(t_i, \mathcal{Y}) = (t_i, \mu)$. If we can find, by inspection, a linear combination of the (t_j, μ) s. t. $(\xi_i, \mu) = \sum_j c_{ij} (t_j, \mu)$, then $\sum_j c_{ij} (t_j, \mathcal{Y})$ is the b.l.u.e. of (ξ_i, μ), and hence is the i^{th} coordinate of P_Ω w.r.t. the ξ's.

Error space and Estimation space

The space of b.l.u.e.s is often called the estimation space and the space of linear estimators with expectation zero is called the error space. Since $(h, \mathcal{Y})$ is a b.l.u.e. iff $h \in \Omega$ and $E(h, \mathcal{Y}) = (h, \mu) = 0 \ \forall \mu \in \Omega$ iff $h \in \Omega^+$, the est'n space is isomorphic to Ω, the error space to Ω^+.

2. The Normality Assumption

The one-dim. normal dist'n has density $f_X(x) = c\, e^{-1/2x^2}$ w.r.t. lebesque measure on R'. It's c.f. $\phi_X(s) = E\, e^{isX} = c \int e^{-1/2(x^2-zisx)}\, dx$.

But $x^2 - zisx = (x-is)^2 + s^2$

So $\underline{\phi_X(s)} = c \int e^{-1/2(x-is)^2}\, dx \cdot e^{-1/2s^2} = \underline{e^{-1/2s^2}}$.

If $Y = (y_1, \cdots, y_n)$ and y_i are indep., normal $(0,1)$, the density of Y, $f_Y(y) = \prod_i c\, e^{-1/2 y_i^2} = c^n e^{-1/2 \Sigma y_i^2}$.
And the c.f. of Y, $\phi_Y(t) = E\, e^{i(t,\, Y)} = \int c^n e^{-1/2\{\Sigma y_i^2 - 2i\Sigma t_i y_i\}} dy$,
$t \in R^n$. But $\Sigma y_i^2 - 2i\Sigma t_i y_i = \sum_j (y_j - it_j)^2 + t_j^2$
So $\phi_Y(t) = e^{-1/2\Sigma t_j^2} = e^{-1/2(t,t)}$.

Let $\mathcal{Y}$ be a random point of V, having c.f. $\phi_{\mathcal{Y}}(t) = E\, e^{i(t, \mathcal{Y})}$
$= e^{-1/2(t,t)\sigma^2}$, $t \in V$ (normal, mean 0, Cov σ^2 I).
Then the <u>family of normal dist'ns</u> is the family of dist'ns obtained by subjecting $\mathcal{Y}$ to linear maps and translations.
If $Z = A\mathcal{Y} + \eta$, $A: V \to W$, $\eta \in W$ the c.f. of $Z = \phi_Z(s) = E\, e^{i(s,Z)}$, $s \in W$.
But $(s,Z) = (s, A\mathcal{Y} + \eta) = (A's, \mathcal{Y}) + (s, \eta)$.
So $\phi_Z(s) = e^{i(s,\,\eta)}\, E\, e^{i(A's, \mathcal{Y})} = e^{i(s,\,\eta)}\, \phi_{\mathcal{Y}}(A's)$
$= e^{-1/2(A's,\, A's)\,\sigma^2 + i(s,\,\eta)} = e^{-1/2(s,\, AA's)\,\sigma^2 + i(s,\,\eta)}$.
But $E\, Z = \eta$ and cov $Z = AA'\sigma^2$
So, the general normal dist'n on W with cov $AA'\sigma^2$, mean η has c.f.

$e^{-1/2(s,\, AA'\sigma^2 s) + i(s,\,\eta)}$, $s \in W$.

<u>Lemma</u>:

If $\mathcal{Y}$ is normal $(\mu, \sigma^2 I)$ in V, and S is a subspace of V, $P_S\mathcal{Y}$ is normal $(P_S\mu, \sigma^2 P_S)$ and hence is <u>spherically normal</u> in S.

Proof:

$$E\, P_S \mathcal{Y} = P_S \mu$$

$\text{Cov}\ P_S \mathcal{Y} = P_S\, \sigma^2\, I\, P_S' = \sigma^2\, P_S.$

Now $P_S \mid S = I$ ($\mid$ denotes the restriction of the map to the subspace S).

And Im $P_S = S$. So $P_S \mathcal{Y}$ is dist'd wpr 1 in S, and is spherically normal there.

The normality assumption in the general linear model is that $\mathcal{Y}$ is spherically normal $(\mu, \sigma^2 I)$ in V, with c.f. $e^{-1/2(t,\ t)\sigma^2 + i(t,\ \mu)}$, $t \in V$.

3. Testing Hypotheses

To test the hypothesis H that the mean is w, some specified subspace of Ω, we decompose $\Omega = w \oplus \Omega - w$, where $\Omega - w$ denotes the orthogonal complement of w, relative to Ω. Then the whole space $V = \Omega^{\perp} \oplus \Omega - w \oplus w$. The F test of H consists in comparing $\|P_{\Omega-w} \mathcal{Y}\|^2$ to $\|P_{\Omega^{\perp}} \mathcal{Y}\|^2$ and rejecting H if the ratio is too large.

$\|P_{\Omega^{\perp}} \mathcal{Y}\|^2$, the squared length of the projection of $\mathcal{Y}$ on $\Omega^{\perp}$, is called the residual sum of squares.

Now $P_S \mathcal{Y}$ is spherically normal in S, a subspace of V, under the normality assumption on $\mathcal{Y}$. And $\|P_S \mathcal{Y}\|^2$ is the length of $P_S \mathcal{Y}$.

Definition:

Let X_i be independent normal $(0, \mu_i)$ r.v.s., $i=1, \cdots, k$.

Then $\sum_{i=1}^{k} X_i$ is distributed as a $\chi^2_{k,\lambda}$ distribution with k degrees of freedom and <u>non-centrality</u> $\lambda = \sum_{i=1}^{k} \mu_i^2$.

<u>Lemma</u>:

Under the normality assumptions, $\|P_S \mathcal{Y}\|^2$ has $\sigma^2 \cdot \chi^2$ dist'n with degrees of freedom the dimension of S and non-centrality $\frac{1}{\sigma^2} \|P_S\, \mu\|^2$.

<u>Proof</u>:

Choose an orthonormal basis for S, say $\xi_1, \cdots, \xi_k$, $k = \dim.\ S$.

Then $P_S \mathcal{Y} = \sum_{i=1}^{k} (\mathcal{Y}, \xi_i)\, \xi_i$ and $\|P_S \mathcal{Y}\|^2 = \sigma^2 \sum_{i=1}^{k} \frac{1}{\sigma^2} (\mathcal{Y}, \xi_i)^2$.

But $(\mathcal{Y}, \xi_i)$ is normal with mean (μ, ξ_i), Var. $\|\xi_i\|^2 \sigma^2 = \sigma^2$ and uncorrelated with (independent of) $(\mathcal{Y}, \xi_j)$, for $j \neq i$. For, $(\xi_i, \sigma^2 I\, \xi_j) = \sigma^2 \delta_{ij}$. And $\frac{1}{\sigma} (\mathcal{Y}, \xi_i)$ is normal with mean $\frac{1}{\sigma} (\mu, \xi_i)$ and var. 1.

So $\|P_S \mathcal{Y}\|^2$ has $\sigma^2 \chi^2_{k,\lambda}$ dist'n., with $\lambda = \sum_{i=1}^{k} \frac{1}{\sigma^2} (\mu, \xi_i)^2 = \frac{1}{\sigma^2} \sum_1^k (\mu, \xi_i)^2$

$= \frac{1}{\sigma^2} \|P_S\, \mu\|^2$. Now, if $S_1 \perp S_2$, where S_1, S_2 are subspaces of V,

$P_{S_1} \mathcal{Y}$ is uncorrelated with (hence independent of) $P_{S_2} \mathcal{Y}$. (See previous corollary). Thus, we conclude that

$$\frac{1}{\dim.\ \Omega - w} \|P_{\Omega - w} \quad\|^2 \Big/ \frac{1}{\dim.\ \Omega^\perp} \|P_{\Omega^\perp} \mathcal{Y}\|^2$$

is the ratio of independent χ^2 variables, divided by their degrees of freedom, and s.t. the numerator has non-centrality $\frac{1}{\sigma^2} \|P_{\Omega - w}\mu\|^2$ (which is zero under H, for then $\mu \in w$), the denominator has non-centrality $\|P_{\Omega^\perp} \mu\|^2 = 0$.

An easy way to find EMS

Notice that $E \| P_S \mathcal{Y} \|^2 = \sum_{i=1}^{k} E(\mathcal{Y}, \xi_i)^2$

But $E(\mathcal{Y}, \xi_i)^2 = \mathrm{Var}(\mathcal{Y}, \xi_i) + E^2(\mathcal{Y}, \xi_i)$

So above $= k\,\sigma^2 + \|P_S \mu\|^2$.

This gives an easy interpretation and method of finding expected mean squares.

$$(1) \qquad \mathrm{EMS} = \frac{E\,\|P_S \mathcal{Y}\|^2}{\dim.\ S} = \sigma^2 + \frac{1}{\dim.\ S}\,\|P_S \mu\|^2$$

4. The Usual Matrix Formulation on R^n

Now we'd like to explain the connection of the previous material with the usual notation.

When we write $Y_{n\times 1} = X_{n\times p}\,\theta_{p\times 1} + \epsilon_{n\times 1}$, $\epsilon_{n\times 1} \sim N(0, \sigma^2 I)$, $\theta \in \Theta = R^p$, we picture a random point $\mathcal{Y}$, distributed in an abstract real inner product space V of dimensional n, the observation space (we can visualize it as 3-space). Y is then the random vector of the coordinates of Y with respect to some orthonormal ordered basis $B = (\xi_1, \cdots, \xi_n)$ of V and $Y_i = (\xi_i, \mathcal{Y})$. The standard basis vector in R^n with 1 in the i^{th} place, 0 elsewhere is then the coordinate vector of ξ_i w.r.t. the ξ's.

$E_\theta\, Y = X\,\theta$ means that $\mathcal{Y}$ is distributed about a mean point μ, which lies in Ω, the mean space (visualize it as a plane through the origin in 3 space), as θ varies through the parameter space R^p. Interpreting $X\,\theta$ as $\sum_{i=1}^{p} \theta_i X_i$, a linear combination of the columns, x_i, of X, we see that Ω is described in a particular way. The columns of X, $x_1, \cdots, x_p$, are the vectors of coordinates w.r.t. B of points $\eta_1, \cdots, \eta_p$ in Ω. Thus, θ is the vector of coordinates of the mean pt. μ w.r.t. the spanning set $\eta_1, \cdots, \eta_p$ of Ω.

$\epsilon \sim \mathcal{N}\,(0,\ \sigma^2\ I)$ means that the random point $\mathcal{Y}$ is distributed in a spherically normal way about its mean point μ.

$$f_\theta(Y) = \frac{1}{(2\pi)^{n/2}}\ e^{-1/2\sigma^2\ \|Y - X\theta\|^2}$$

We can picture a dist'n. of mass about μ in 3 space s.t. the greatest density of mass is at μ, and the density is constant on spheres about μ, decreasing exponentially with the increasing radius of the sphere.

Given an observed value of Y, y, we set up the normal equations, $X'(y - X\theta) = 0$, and seek a solution $\hat{\theta}$.

Interpreting matrix multiplication of X' and $y - X\theta$ as the values of the inner product of $(y - X\theta)$ with each row of X', i.e., with each column of X, the $\hat{\theta}$ we seek is the set of coordinates w.r.t. $\eta_1, \cdots, \eta_p$ of that point $\mu^* \in \Omega$ s.t. $y - \mu^*$ is orthogonal to $\eta_1, \cdots, \eta_p$ and hence to Ω, since the η's span Ω.

But the unique such $\mu^* = P_\Omega$

Note that although μ^* is unique, $\hat{\theta}$, its coordinates w.r.t. the η's are unique iff the η's are linearly independent (X has full rank). If the η's aren't l. indep., any vector in Ω can be expressed as a linear combination of the η's in many ways.

The next step in the usual computations is $X'X\hat{\theta} = X'y$. Now $X'X$ is the matrix representation of a map on R^p to R^p of rank $r \leq p$. For,

$$\begin{tikzcd} \theta \arrow[r, shift left, "X"] & \Omega \arrow[l, shift left, "X'"] \end{tikzcd}$$

Obviously $\text{Ker } X \subset \text{Ker } X'X$.

But $(X'X) = (X'X)'$.

Hence $\text{Ker } X'X = (\text{Im } X'X)^{\perp}$.

Now $\text{Im } X'X \subset \text{Im } X'$, clearly.

So $(\text{Im } X'X)^{\perp} \supset (\text{Im } X')^{\perp} = \text{Ker } X$.

Thus, $\text{Ker } X'X \supset \text{Ker } X$ and equality holds.

By a dimensionality argument, since both maps have the same domain, X and $X'X$ have the same rank.

When $r < p$, $X'X$ has a Kernel and is constant on the cosets $\theta \oplus \operatorname{Ker} X'X$, $\theta \in \Theta$. A <u>generalized</u> <u>inverse</u>, G, of $X'X$, i.e., a matrix s.t. $(X'X)\, G\, (X'X) = X'X$, can be regarded as the representation of a map which will, for each point p in the image of $X'X$, pick out a point in the domain which maps into p, i.e., a point in the proper coset, in a linear way.

$$\Theta \xrightarrow{X'X} \Theta^p \xrightarrow{G} \Theta \xrightarrow{X'X} \Theta^p$$

This follows by the interpretation of matrix multiplication as composition of linear maps.

Thus, $\hat{\theta} = G\, X'y$ yields a point in $(X'X)^{-1}\{X'y\}$.

It is not an accident that $\hat{\theta}$ consists of p linear combinations of $X'y$. Interpreting matrix multiplication as before, $X'y$ is the inner product of y with each of the columns of X. It is the values at the vector $\mathcal{Y} \in V$ of the p linear functionals on Ω, $(\eta_1, \cdot), \cdots, (\eta_p, \cdot)$, which (when composed with $\mathcal{Y}$) span the estimation space.

Finally, by the Gauss-Markov theorem, if $\theta \to (\underline{f}, \theta)$ is an est'le l.p.f., its b.l.u.e. is $(\underline{f}, \hat{\theta})$, which doesn't depend on which G we used to calculate $\hat{\theta}$.

5. Note on Side Conditions on the Parameters

If X is not a 1 - 1 map, the least-squares est'r $\hat{\theta}(y)$ is not unique (although the derived b.l.u.e.s of est'le l.p.f.s are). Suppose, we would like to place side conditions on $\hat{\theta}(y)$ to single out a particular point in $X^{-1}\{P_\Omega y\}$.

Picture:

$$
\begin{tikzcd}
R^{p-r} & & \Omega \subset V \\
& \Theta \arrow[ul, "H"] \arrow[ur, "X"'] &
\end{tikzcd}
$$

Ω: r dim. Θ: p dim. $r < p$

Choose an H with Ker H $\cap$ Ker X = 0 and dim. Ker H = r. Since p = dim. Θ = dim. Im H + dim. Ker H, and dim Im H $\leq$ p - r, dim. Im H = p - r, and H is full rank.

That is, we are choosing p - r linearly independent linear functionals h_i on Θ s.t. Ker H = $\bigcap_{i=1}^{p-r}$ Ker h_i doesn't intersect Ker X.

Then Θ = Ker H $\oplus$ Ker X by a dimensionality argument. Hence, given any point $\mu \in \Omega$, the coset x $\oplus$ Ker X in Θ space mapping into it $(X^{-1}\{\mu\})$ intersects Ker H in exactly one point, namely, projection on Ker H along Ker X of the coset x $\oplus$ Ker X. So the point θ in $X^{-1}\{P_\Omega y\}$, satisfying $H\theta = 0$ is uniquely determined.

6. Note on Replication

Let x be a point in R^n with coordinates $(x_1, \cdots, x_n)$ w.r.t. the standard basis. Let S be the subspace of R^n, where the coordinates are equal in r batches, $S = \{x: x_1 = \cdots = x_{n_1}, x_{n_1+1} = \cdots = x_{n_1+n_2}, \cdots$

$x_{\sum_1^{r-1} n_i+1} = \cdots = x_{\sum_1^{r} n_i} \}$, $\sum_1^r n_i = n$.

We seek $P_S x$, the projection of x on S. We claim $P = \left(\frac{\sum_1^{n_1} x_i}{n_1}, \cdots, \frac{\sum_{n_1+1}^{n_1+n_2} x_i}{n_2}, \right.$

$\cdots$ etc.$) = P_S x$. For, clearly $p \in S$.

And $(x - p, s) = (x, s) - (p, s)$ for any $s \in S$.

Let $s = (s_1, \cdots, s_1, s_2, \cdots, s_2, \cdots)$.

$$\text{Then } (x, s) = \sum_{i=1}^{r} s_i \sum_{j=\sum_{k=1}^{i-1} n_k+1}^{\sum_1^i n_k} = (p, s)$$

So $x - p \in S^{\perp}$ and $p = P_s x$.

Examples

The one-way layout

$E\ y_{ij} = \alpha_i \quad i = 1, \cdots, I,\ j = 1, \cdots, n_i$

$\sum_{i=1}^{I} n_i = n$

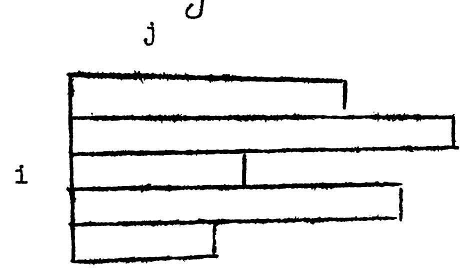
(From now on, $\mathcal{Y}_{ijk} \cdots$ will denote a random variable, the linear functional $(\xi_{ijk} \cdots, \mathcal{Y})$, where $\xi_{ijk} \cdots$ is the vector in R^n placing a 1 in the ijk ••• th cell, 0's elsewhere.)

dim V = n, dim Ω = I.

1. Under Ω, the B.L.U.E.s are spanned by $y_{i.}$, $i = 1, \cdots, I$.

$(y_{i.} = \sum_{j=1}^{n_i} y_{ij})$

For, we can picture putting 1 in the i^{th} section and 0's elsewhere. The resulting vectors span Ω, hence their dots with $\mathcal{Y}$ span the estimation space.



Now $E\ \bar{y}_{i.} = \alpha_i$

$= i,j^{th}$ coordinate of μ under $\Omega = (\xi_{ij}, \mu)$.

So $\bar{y}_{i.}$ is the i,j^{th} coord. of $P_\Omega \mathcal{Y} = (\xi_{ij}, P_\Omega \mathcal{Y})$.

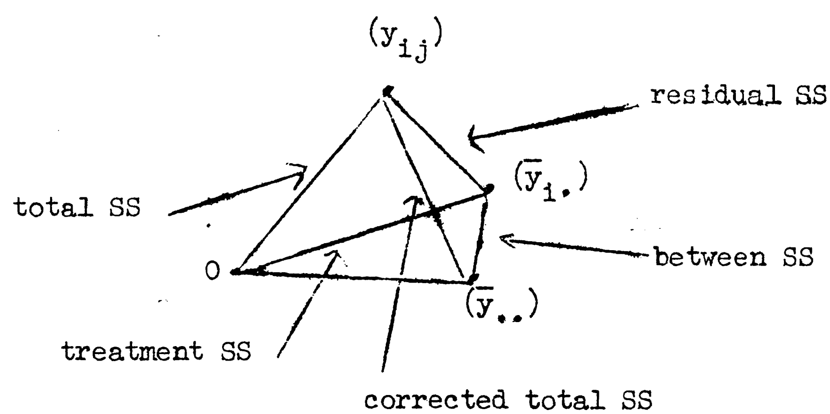
2. Let Ω_1 = the subspace of Ω, where all coordinates are equal. dim $\Omega_1 = 1$. It corresponds to the hypothesis $\alpha_i = \alpha_{i'}$. Under Ω_1, the b.l.u.e.s are spanned by $y_{\cdot\cdot}$.

Now $E\,\bar{y}_{\cdot\cdot} = \bar{\alpha}_{\cdot} = (\xi_{ij}, \mu)$, $\mu \in \Omega_1$.

So $\bar{y}_{\cdot\cdot} = (\xi_{ij}, P_{\Omega_1} y)$.

Thus, $(\xi_{ij}, P_{\Omega-\Omega_1} y) = \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}$.

Similarly, $(\xi_{ij}, P_{V-\Omega} y) = y_{ij} - \bar{y}_{i\cdot}$.



We will use the convention of denoting the coordinates of a point w.r.t. ξ_{ij} by $(\cdot)$. Then, in the above picture, the squared length of a line is the corresponding sum of squares.

Calculations

Compute $\|\mathcal{Y}\|^2 = \|(y_{ij})\|^2 = \sum_{ij} y_{ij}$ = total SS and

$\|P_{\Omega_1}\mathcal{Y}\|^2 = \|(\bar{y}_{..})\|^2 = \sum_{ij} \bar{y}^2 = \dfrac{y^2_{..}}{n}$ = correction for the mean.

Then $\|P_{V-\Omega_1}\mathcal{Y}\|^2 = \|(y_{ij} - \bar{y}_{..})\|^2 = \|\mathcal{Y}\|^2 - \|P_{\Omega_1}\mathcal{Y}\|^2$ = total SS - cor. for mean

= corrected total SS, for $V = \Omega_1 \oplus V - \Omega_1$.

Compute $\|P_{\Omega}\mathcal{Y}\|^2 = \|(\bar{y}_{i.})\|^2 = \sum_{ij} \bar{y}^2_{i.} = \sum_{i} \dfrac{y^2_{i.}}{n_i}$ = treatment SS.

Then $\|P_{\Omega-\Omega_1}\mathcal{Y}\|^2 = \|(\bar{y}_{i.} - \bar{y}_{..})\|^2 = \|P_{\Omega}\mathcal{Y}\|^2 - \|P_{\Omega_1}\mathcal{Y}\|^2$ = treatment SS

- cor. for mean = Between treat. SS, for $\Omega = \Omega_1 \oplus \Omega - \Omega_1$.

Finally, $\|P_{V-\Omega}\mathcal{Y}\|^2 = \|P_{V-\Omega_1}\mathcal{Y}\|^2 - \|P_{\Omega-\Omega_1}\mathcal{Y}\|^2 = \|(y_{ij} - \bar{y}_{i.})\|^2$

= total SS - between treatments SS.= within treatments SS = residual SS,

for $V - \Omega_1 = V - \Omega \oplus \Omega - \Omega_1$.

To compute expected values of SS:

$E\,\|P_{V-\Omega}\mathcal{Y}\|^2 = \sigma^2 \dim V - \Omega = (N - I)\,\sigma^2$

E Between treats. SS = $E\,\|P_{\Omega-\Omega_1}\mathcal{Y}\|^2 = \sigma^2 \dim \Omega - \Omega_1 + \|P_{\Omega-\Omega_1}\mu\|^2$

But $\dim \Omega - \Omega_1 = \dim \Omega - \dim \Omega_1 = I - 1$.

And $\|P_{\Omega-\Omega_1}\mu\|^2 = \|P_{\Omega}\mu\|^2 - \|P_{\Omega_1}\mu\|^2$, as $\Omega = \Omega_1 \oplus \Omega - \Omega_1$.

Now $P_{\Omega_1}\mu = \frac{\sum_{ij} \alpha_i}{n} = \sum_i \frac{n_i}{n} \alpha_i = \bar{\alpha}_w$, a weighted average of the α's.

So $\|P_{\Omega-\Omega_1}\mu\|^2 = \|\mu\|^2 - \|P_{\Omega_1}\mu\|^2 = \sum_{ij} \alpha_i^2 - \frac{(\sum_{ij} \alpha_i)^2}{n} = \sum_i n_i \alpha_i^2 - \frac{(\sum_{ij} \alpha_i)^2}{n}$

$= \sum_i n_i(\alpha_i - \bar{\alpha}_w)^2$

Every α_i is est'le, since $\alpha_i = (\xi_{ij}, \mu)$, $\mu \in \Omega$.

The b.l.u.e. of α_i is $\bar{y}_{i.}$, since $E\,\bar{y}_{i.} = \alpha_i$ and $\bar{y}_{i.}$ is a b.l.u.e.

The two-way layout, equal nos.

$$E\; y_{ijk} \cdots = \alpha_i + \beta_j + \gamma_{ij}, \quad i = 1, \cdots, I,\ j = 1, \cdots, J$$
$$k = 1, \cdots, K.$$

1. Under Ω, the b.l.u.e.s are spanned by $y_{ij.}$, for their coefficient vectors span Ω (this is always the case when there is replication).

Now $E\,\bar{y}_{ij.} = \alpha_i + \beta_j + \gamma_{ij} = (\xi_{ijk}, \mu)$.

So $(\xi_{ijk}, P_\Omega \mathcal{Y}) = \bar{y}_{ij.}$.

$\dim \Omega = IJ$.

2. Consider the hypothesis of no interaction $(\gamma_{ij} = \gamma_{i'j'})$. Then the mean space is restricted to $\Omega_\gamma \subset \Omega$, where $\Omega_\gamma = \{(x_{ijk}): \ x_{ijk} = a_i + b_j$ for some a_i, b_j real$\}$. $\dim \Omega_\gamma = I + J - 1$, for the kernel of the parametrization is the subspace where all the α's $= -$ the β's (dim 1).

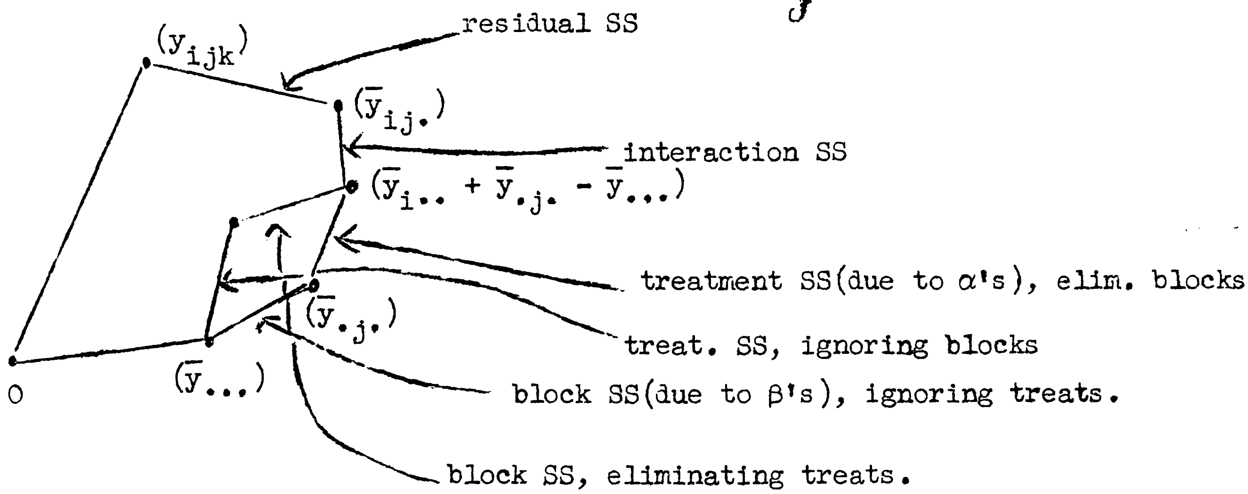
Under Ω_γ, the b.l.u.e.s are spanned by $y_{i..}$, $y_{.j.}$.

Now $E\,\bar{y}_{i..} = \alpha_i + \bar{\beta}_. + \bar{\gamma}_{..}$, $E\,\bar{y}_{.j.} = \bar{\alpha}_. + \beta_j + \bar{\gamma}_{..}$, and $E\,\bar{y}_{...}$

$$= E\,\bar{y}_{...} = E \sum_{i=1}^{I} \frac{\bar{y}_{i..}}{I} = \bar{\alpha}_. + \bar{\beta}_. + \bar{\gamma}_{..}.$$

Now $\bar{y}_{i..} + \bar{y}_{.j.} - \bar{y}_{...}$ is a linear combination of b.l.u.e.s with the correct expected value, namely, $\alpha_i + \beta_j + \bar{\gamma}_{..} = (\xi_{ijk}, \mu)$, $\mu \in \Omega_\gamma$. Hence $(\bar{y}_{i..} + \bar{y}_{ij.} - \bar{y}_{...}) = P_{\Omega_\gamma} y)$.

3. Next consider the hypothesis of no interaction and no treatment effects (γ's equal, α's equal) with mean space $\Omega_\beta \subset \Omega_\gamma$.
$\Omega_\beta = \{(x_{ijk}): x_{ijk} = x_{i'jk'}\}$. $\dim \Omega_\beta = J$.
The b.l.u.e.s are then spanned by $y_{.j.}$.
Now $E\,\bar{y}_{.j.} = \bar{\alpha}_. + \beta_j + \bar{\gamma}_{..} = (\xi_{ijk}, \mu)$, $\mu \in \Omega_\beta$. So $(\bar{y}_{.j.}) = P_{\Omega_2} y)$.
4. Finally, consider the hypothesis of no effects with mean space 1. dim 1 = 1. Then the b.l.u.e.s are spanned by $y_{...}$. And $E\,\bar{y}_{...} = \bar{\alpha}_. + \bar{\beta}_. + \bar{\gamma}_{..} = (\xi_{ijk}, \mu)$, $\mu \in 1$. So $(\bar{y}_{...}) = P_1 y$.



$\Omega_\alpha = \{(x_{ijk}): \ x_{ijk} = x_{ij'k'}\}$ (γ's equal, β's equal).

Note that $\Omega_\alpha \cap \Omega_\beta = 1$. And $\Omega_\alpha \oplus \Omega_\beta = \Omega_\gamma$.

So $\Omega_\gamma = 1 \oplus \Omega_\alpha - 1 \oplus \Omega_\beta - 1$.

This is <u>not</u> an orthogonal decomposition.

$$\|P_1 \mathcal{Y}\|^2 + \|P_{\Omega_\alpha - 1} \mathcal{Y}\|^2 + \|P_{\Omega_\beta - 1} \mathcal{Y}\|^2 = \|P_{\Omega_\gamma} \mathcal{Y}\|^2$$

cor. for mean   treat.(ign. blocks)   blocks(ign. treats)   no interaction

iff $\mathcal{Y}_\alpha = P_{\Omega_\alpha - 1}\mathcal{Y} \perp P_{\Omega_\beta - 1}\mathcal{Y} = \mathcal{Y}_\beta$. Let $z_\alpha = P_{\Omega_\alpha}\mathcal{Y}$, $z_\beta = P_{\Omega_\beta}\mathcal{Y}$.

Then $z_\alpha = \mathcal{Y}_\alpha + c1$, $z_\beta = \mathcal{Y}_\beta + c1$, for some c, where $c1 = P_1\mathcal{Y} = P_1 z_\alpha$

$= P_1 z_\beta$.

$(y_\alpha, y_\beta) = (z_\alpha - c1, z_\beta - c1) = (z_\alpha, z_\beta) - (z_\alpha, c1) - (z_\beta, c1) + (c1, c1)$

$= (z_\alpha, z_\beta) - (c1, c1)$, as $(z_\alpha, c1) = (c1, c1)$.

Hence, it is zero iff $\sum_{ijk} \bar{y}_{i..}\,\bar{y}_{.j.} = \sum_{ijk} (\bar{y}_{...})^2$.

<u>Computations</u>

1. $\|P_{\Omega_\alpha - 1}\mathcal{Y}\|^2 = \|P_{\Omega_\alpha}\mathcal{Y}\|^2 - \|P_1\mathcal{Y}\|^2 = \sum_{ijk} \bar{y}^2_{i..} - \sum_{ijk} \bar{y}^2_{...} = \sum_i \frac{y^2_{i..}}{JK}$

$- \frac{y^2_{...}}{IJK}$ = <u>treatment</u> <u>SS</u>.

2. $\|P_{\Omega_\beta - 1}\mathcal{Y}\|^2 = \|P_{\Omega_\beta}\mathcal{Y}\|^2 - \|P_1\mathcal{Y}\|^2 = \sum_j \frac{y^2_{\cdot j \cdot}}{IK} - \frac{y^2_{\cdot\cdot\cdot}}{IJK}$ = block SS.

3. $\|P_{V-1}\mathcal{Y}\|^2 = \|\mathcal{Y}\|^2 - \|P_1\mathcal{Y}\|^2 = \sum_{ijk} y^2_{ijk} - \frac{y^2_{\cdot\cdot\cdot}}{IJK}$ = cor. total SS.

4. $\|P_{\Omega-1}\mathcal{Y}\|^2 = \|P_\Omega\mathcal{Y}\|^2 - \|P_1\mathcal{Y}\|^2 = \sum_{ijk} \bar{y}^2_{ij\cdot} - \frac{y^2_{\cdot\cdot\cdot}}{IJK} = \sum_{ij} \frac{y^2_{ij\cdot}}{K}$

$- \frac{y^2_{\cdot\cdot\cdot}}{IJK}$ = cor. cell SS.

5. $\|P_{V-\Omega}\mathcal{Y}\|^2 = \|P_{V-1}\mathcal{Y}\|^2 - \|P_{\Omega-1}\mathcal{Y}\|^2$ = cor. total SS - cor. cell SS

= residual SS.

6. $\|P_{\Omega-\Omega_\gamma}\mathcal{Y}\|^2 = \|P_{\Omega-1}\mathcal{Y}\|^2 - \|P_{\Omega_\alpha-1}\mathcal{Y}\|^2 - \|P_{\Omega_\beta-1}\mathcal{Y}\|^2$ = cor. cell SS

- treat. SS - block SS = interaction SS.

Under Ω, none of the α's, β's, or γ's are est'le.

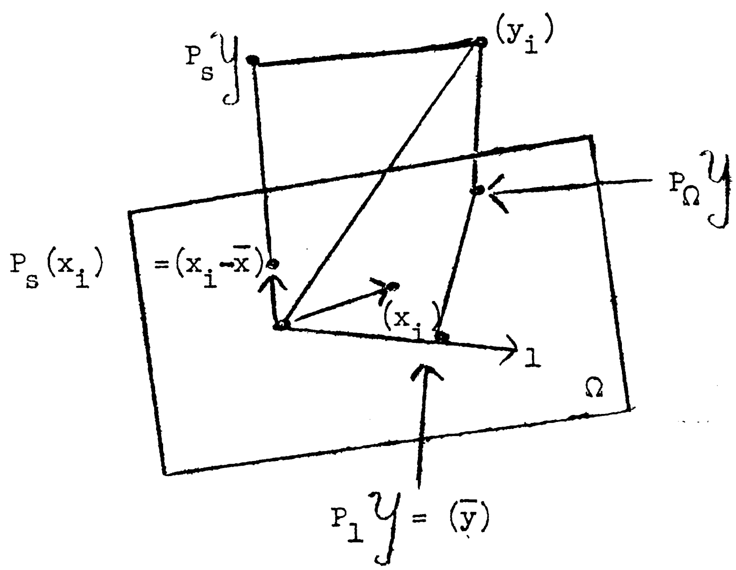
Under Ω_γ (no interaction), any linear contrast in the α's or β's, say

$\Sigma c_i\alpha_i$, where $\Sigma c_i = 0$, is est'le, because $E \,\Sigma c_i\bar{y}_{i\cdot\cdot} = \Sigma c_i(\alpha_i+\beta_\cdot+\bar{\gamma}_{\cdot\cdot})$

$= \Sigma c_i\alpha_i$, and $\bar{y}_{i\cdot\cdot}$ are b.l.u.e.s, under Ω_1.

Simple Linear Regression

$E\, y_i = \alpha + \beta\, x_i \quad , \qquad i = 1, \cdots, n.$



Let 1 = the subspace where all coordinates are equal, corresponding to the hypothesis $\beta = 0$. Let $S = \Omega - 1$.

Then $P_1 \mathcal{Y} = (\bar{y})$ (See note)

And $$P_s \mathcal{Y} = \frac{((x_i - \bar{x}), (y_i))}{\|(x_i - \bar{x})\|} \cdot \frac{(x_i - \bar{x})}{\|(x_i - \bar{x})\|}$$

$$= \frac{\sum_i x_i y_i}{\sum_i (x_i - \bar{x})^2} (x_i - \bar{x}).$$

So $P_\Omega \mathcal{Y} = P_1 \mathcal{Y} + P_s \mathcal{Y}$

(as $\Omega = 1 \oplus S$ and $1 \perp S$).

$$P_\Omega \mathcal{Y} = \left(\bar{y} - \frac{\sum\limits_j x_j y_j}{\sum\limits_j (x_j - \bar{x})^2}\,\bar{x} + \frac{\sum\limits_j x_j y_j}{\sum\limits_j (x_j - \bar{x})^2}\,x_i\right)$$

$$\text{So } \hat{\alpha} = \bar{y} - \frac{\sum\limits_j x_j y_j}{\sum\limits_j (x_j - \bar{x})^2}\,\bar{x}$$

$$\hat{\beta} = \frac{\sum\limits_j x_j y_j}{\sum\limits_j (x_j - \bar{x})^2} \;.$$

$$\text{Now } \quad \|P_{\Omega-1}\mathcal{Y}\|^2 = \frac{((x_i - \bar{x}), (y_i))^2}{\|(x_i - \bar{x})\|^2} = \frac{(\sum\limits_j x_j y_j)^2}{\sum\limits_j (x_j - \bar{x})^2}$$

= SS due to β not being zero = SS <u>due</u> <u>to</u> <u>regression</u>.

$$\|P_1\mathcal{Y}\|^2 = \sum_i \bar{y}^2 = \frac{y_{\cdot}^2}{n} = \underline{\text{cor.}}\ \underline{\text{for}}\ \underline{\text{the}}\ \underline{\text{mean}}$$

$$\|P_{V-1}\mathcal{Y}\|^2 = \sum_i y_i^2 - \frac{y_{\cdot}^2}{n} = \underline{\text{cor.}}\ \underline{\text{total}}\ \underline{\text{SS}}$$

So $\|P_{V-\Omega}\mathcal{Y}\|^2$ = <u>residual</u> <u>SS</u> = $\|P_{V-1}\mathcal{Y}\| - \|P_{\Omega-1}\mathcal{Y}\|^2$ = cor. total SS

- SS due to regression.

References

[1] P.R. Halmos, Finite-Dimensional Vector Spaces, Princeton, Van Nostrand Press, 1958 (2nd ed.).

[2] J. Kiefer, "Notes on Statistical Inference", unpublished.

[3] William Kruskal, "The Coordinate-Free Approach to Gauss-Markov Estimation, and its Applications to Missing and Extra Observations", Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, pp. 435 - 451.